Original scientific paper

UDC: 616-089.884:612.78

Received: April 11, 2025.
Revised: July 13, 2025.
Accepted: July 21, 2025.

10.23947/2334-8496-2025-13-2-289-310

Check for updates

Can Voice Characteristics Predict the Severity of Depression: A Study on Serbian-Speaking Participants

Gordana Calić[1], Branimir Radmanović[2,3], Mirjana Petrović-Lazić[1], Dragana Ignjatović Ristić[2,3], Nikola Subotić[2,3], Milena Mladenović[3,4]*

[1]Department of Speech and Language Pathology, Faculty of Special Education and Rehabilitation, University of Belgrade, Belgrade, Serbia, e-mail: calicgordana@yahoo.com, carica@rcub.bg.ac.rs
[2]Department of Psychiatry, Faculty of Medical Sciences, University of Kragujevac, Kragujevac, Serbia, e-mail: biokg2005@yahoo.com, draganaristic4@gmail.com, nikolasrf@gmail.com
[3] Psychiatric Clinic, University Clinical Center Kragujevac, Kragujevac, Serbia, e-mail: milena.jovicic@uni.kg.ac.rs
[4]Department of Psychology, Faculty of Medical Sciences, University of Kragujevac, Kragujevac, Serbia

Abstract: There is a growing interest in detecting depression through vocal indicators for the purpose of early diagnosis and therapeutic monitoring. Thus, research on voice characteristics in different language areas among individuals with depression may potentially contribute to the standardization of vocal analysis and the development of automatic recognition programs. This study aims to determine whether specific voice characteristics can predict the severity of depression using the Montgomery-Asberg Depression Rating Scale (MADRS) in a sample of Serbian-speaking participants. The analysis included perceptual (GRBAS scale parameters) and acoustic (parameters of frequency variability, intensity variability, and noise and tremor estimation using the MDVP software) voice characteristics in a sample of 100 participants. The sample was divided into two groups: an experimental group of participants diagnosed with depressive disorder (N = 45), including an equal number of participants with mild, moderate, and severe depression (N = 15), and a control group of participants without a depressive disorder diagnosis or depression symptoms (N = 55). The prediction of depression severity based on voice characteristics was conducted using hierarchical regression analysis. The results indicate statistically significant differences in nearly all acoustic and all perceptual voice characteristics among participants with different levels of depression symptoms (MADRS score). Post-hoc analysis revealed no differences in acoustic characteristics between subgroups with different depression severity levels. However, significant differences in perceptual characteristics were found among all subgroups, except between mild and moderate depression. After controlling for gender, age, and smoking status, depression severity demonstrated statistically significant effects on nearly all acoustic and all perceptual voice characteristics. Both perceptual and acoustic voice characteristics can predict the severity of depression. The acoustic parameter of peak amplitude variation (vAm) and the perceptual parameters of hoarseness (G), breathiness (B), asthenia (A), and strain (S) were significant predictors of depression severity. Voice may hold potential as an indicative marker in predicting the severity of depression measured by the MADRS scale. The acoustic parameter related to intensity variation and the perceptual parameters of the GRBAS scale (except voice roughness) appear to be promising voice characteristics in training depression recognition models. Identifying vocal indicators as markers for detecting mental disorders, such as depression, through regression analysis may serve as a foundation for the development of artificial intelligence models for its recognition and may have future clinical relevance.

Keywords: *depression severity, predictors, regression, Serbian language, acoustic analysis, perceptual analysis, biomarker, depression recognition.*

Introduction

Research efforts to detect and monitor mental disorders, such as major depressive disorder (hereafter depression), through objective biomarkers, such as voice, have been growing in recent years. While there is an increasing number of studies exploring vocal characteristics in depression and various high-precision classification models, a voice biomarker for its detection has not been validated yet. Therefore, patient self-reporting remains the only available diagnostic resource (Zhang et al., 2020), alongside sig-

*Corresponding author: calicgordana@yahoo.com

© 2025 by the authors. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

nificant expertise from professionals (Huang et al., 2024).

In addition to providing greater objectivity and facilitating the diagnostic process, voice-based depression recognition models offer the possibility of collecting data in a relatively easy and non-intrusive manner, while the recording procedure does not require high costs (Huang et al., 2024). However, the models vary in recognition accuracy due to different parameters analyzed in studies, speech tasks, assessment scales, methods of analysis, sample heterogeneity, etc. Common machine-learning approaches for depression recognition include linear regression (Mundt et al., 2012; Silva et al., 2024; Zhao et al., 2022; Yang et al., 2013; Wadle et al., 2024), a support vector machine (Kiss and Jenei, 2020; Liu et al., 2016; Menne et al., 2024; Sahu and Espy-Wilson, 2016; Yalamanchili et al., 2020; Williamson et al., 2018), a Gaussian mixture model (Afshan et al., 2018; Cummins et al., 2015), a combination of methods (Alghowinem et al., 2013; Jiang et al., 2017; Shin et al., 2021) or neural networks (Chlasta et al., 2019; Liang et al., 2024; Rejaibi et al., 2022; Seneviratne and Espy-Wilson, 2021; Wang et al., 2023).

It is observed that the existing literature on this topic is based on studies conducted predominantly in Western and increasingly Eastern countries, which highlights the need for further studies in other language areas to verify the linguistic and cross-cultural consistency of the parameters. Unlike English, where the accented syllable is usually characterized by a higher fundamental frequency (F0), longer duration and greater intensity (stress-accented language), an accented syllable in Serbian is characterized by a change in pitch and duration (pitch-accented language), but not a change in intensity compared to an unaccented syllable or different types of accents (Bjelica, 2012). Also, accent can be on any syllable, except the last one, unlike e.g. Czech and Polish, which, like Serbian, belong to the Slavic languages, and where the accent is always tied to a certain position in the word. In the Serbian language, the tonic accent is phonemic, that is, changes in the pitch of an accented syllable can change the meaning of a word. In contrast to most Slavic languages, Serbian prosodic system is characterized by a combination of tonal and quantitative accent, where tone pitch (ascending/descending) and vowel length (short/long) are phonologically relevant and together participate in distinguishing meaning (Bjelica, 2012). In Serbian, unlike, for example, English, vowels in unstressed syllables remain of the same vocal quality (without reduction) (Nikolić, 2016) which could have impact on differences in prosodic structure. Eastern languages, such as Mandarin, are mostly tonal languages, meaning that each syllable has a specific tone and changing the tone also changes the meaning (lexical function) (Yu et al., 2017). Differences in accentuation between languages can affect speech production and thus vocal biomarkers, such as parameters that express changes in the F0 of the voice and its variability. Therefore, it is also important to take into account the prosodic specificities of a particular language when analyzing voice parameters in the context of emotional states, such as depression. Additionally, research samples often neglect participants with mild depression and include unequal numbers of participants with moderate and severe depression, which limits the prediction. Existing studies in the Serbian-speaking area mostly focus on identifying differences between participants with depression and a control group, while insufficient attention has been given to developing models that enable reliable depression prediction.

Previous paper (Calić et al., 2022a) focused on the discriminative role of voice characteristics in distinguishing between groups with and without depression, while this study explores their predictive role. We included additional voice characteristics, both acoustic and perceptual, in accordance with recommendations from authors in this field to incorporate parameters from different domains. Although research studies most commonly use the Hamilton Depression Rating Scale (HAM-D) and the Beck Depression Inventory (BDI, BDI-II), we used the Montgomery-Asberg Depression Rating Scale (MADRS) due to its good validity and higher discriminative power for moderate and severe depression compared to HAM-D (Müller et al., 2003), as well as its more accurate discrimination of individuals without depression symptoms within primary healthcare compared to BDI-II (Nejati et al., 2020). In addition, the sample included an equal number of participants with different levels of depression severity.

To our knowledge, this study represents the first attempt to identify depression severity predictors based on voice characteristics in the Serbian-speaking area.

Mechanisms Underlying Voice and Depression

Reviewing the literature revealed several potential mechanisms underlying altered voice characteristics in depression. They can be classified into three general groups: neurophysiological/neurobiological, cognitive/psychological, and socio-emotional.

Some authors emphasize neurophysiological mechanisms, such as the impact of psychomotor impairment (slowing of thoughts and limited movements), as a dominant symptom in depression, on speech and voice. Psychomotor slowing is thought to affect laryngeal dynamics and control (Quatieri and Malyska, 2012), and authors most often associate this factor with the voice characteristics that indicate precision in motor control during vocal production, such as voice quality features (Jitter, Shimmer, etc.) (Quatieri and Malyska, 2012; Zhang et al., 2020) and also prosodic (like pitch variability, speech rate and pause time) (Cannizzaro et al., 2004). Changes in muscle tone of the vocal tract as well as the respiratory system, often associated with fatigue in depression (due to changes in the autonomic nervous system), can affect the voice (Zhao et al., 2022). The role of dopamine (DA) deficiency has been emphasized in some studies (Darby et al., 1984), while others point to the contribution of serotonin (5-HT) (Zhao et al., 2022) as a potential neurobiological mechanism underlying altered voice characteristics. These neurotransmitter imbalances are believed to affect neural circuits involving the prefrontal cortex and basal ganglia (Vahid-Ansari and Albert, 2021), which are crucial for motor planning and vocal control, thereby contributing to psychomotor slowing and altered voice production in depression (Yamamoto et al., 2020). In addition to neurophysiological and neurobiological mechanisms, cognitive, psychological and socio-emotional factors also play an important role.

Cognitive deficits, such as impairment of working memory, attention, and executive functions, can affect speech planning and production (Alpert et al., 2001). Cognitive mechanisms are thought to underlie the reduced rate of speech and the greater number of pauses and their longer duration in people with depression. Some authors point out that the total number, duration and variability of pauses in automatic speech tasks (e.g. reading) reflect psychomotor slowing, while cognitive factors are more closely associated with free speech tasks (e.g. word finding during an interview) (Alpert et al., 2001; Mundt et al., 2007). Psychological factors, such as low arousal, lack of motivation, and anhedonia have also been proposed as contributing factors (Almaghrabi et al., 2023).

Ellgring and Scherer (1996) point out that if psychomotor impairment resulting from neurological dysfunction (like neurotransmitter deficiency) were the cause, there would be a general effect of muscle rigidity on speech production, as well as the influence of cognitive deficits, and no, for example, gender differences in voice characteristics among people with depression. They highlight the socio-emotional hypothesis, suggesting that different patterns of speech and voice quality are determined by the type of underlying emotion. Accordingly, if the underlying state is apathy, one would expect lower F0, a slower speech rate, and longer pauses, whereas anxiety is expected to show the opposite pattern. It is assumed that psychomotor slowing is primarily associated with sadness, whereas agitation may reflect a combination of sadness and anxiety (Alpert et al., 2001).

Given the methodological differences across studies, the complex nature of voice, and the heterogeneity of factors associated with depression, the specific underlying mechanism remains an open question. Although the analyses of voice characteristics in depression cannot directly identify the underlying causes, they may enhance understanding of the psychopathological processes involved and inform future research aimed at uncovering these mechanisms.

Voice-Based Depression Recognition

Correlation analyses of voice and depression severity

Numerous research studies confirm the presence of differences in certain voice characteristics, both perceptual (Darby et al., 1984) and more frequently analyzed acoustic ones (Alpert et al., 2001; Jia et al., 2019; Silva et al., 2024; Taguchi et al., 2017; Wang et al., 2019; Zhao et al., 2022), between participants with and without depression. Several studies have also shown that some of these characteristics correlate with the severity of depression (Hönig et al., 2014; Mundt et al., 2012; Yamamoto et al., 2020; Zhao et al., 2022). For example, a Japanese study (Yamamoto et al., 2020) shows that prosodic features

(speech rate, pause period, and response time) significantly correlate with depression severity measured by the Hamilton Depression Rating Scale (HAMD-17). A study conducted in the USA (Cannizzaro et al., 2004) shows that speech rate is significantly negatively correlated with depression severity, but the correlation with percent pause time was not significant. This is contrary to the results of Mundt et al. (2007, 2012) who replicated this finding in a larger sample and demonstrated a significant correlation of both speech rate and percent pause time with depression severity. Sample size and heterogeneity could explain the inconsistency in results. Also, a Chinese study (Zhao et al., 2022) found a positive correlation between spectral parameters, specifically two Mel-frequency cepstral coefficients (MFCC4 and MFCC7), and the Patient Health Questionnaire (PHQ-9), while in another study in Japanese sample (Taguchi et al., 2018) the MFCC coefficients were not significantly associated with severity of depression. Although the speech task (reading paragraphs and numbers) in these studies was the same, it is possible that language differences, sample heterogeneity, different scales for assessing the severity of depression and different voice recording methods may account for differences in results. Some studies (Quatieri and Malyska, 2012) show that voice quality parameters, Jitter and Shimmer, correlate with depression severity (HAMD-17), unlike the F0 parameter. On the other hand, other studies (Mundt et al., 2007; 2012; Hönig et al., 2014) found a significant correlation between F0 and F0 variability with depression severity. It is possible that different languages and speech tasks used in these studies led to inconsistent results.

Predictive analysis of depression severity using voice characteristics

While examining the predictive role of voice in recognizing depression, Hashim et al. (2017) used multiple linear regression and found indications of gender differences. Specifically, acoustic voice characteristics based on reading showed significant predictive value for the HAMD score in both genders, while for the BDI-II, this was only true for men. However, according to the authors, the limiting factor of their study could be that it did not include potential confounding variables, such as smoking history and the voice of professional voice users. By also analyzing voice characteristics based on reading but in a Chinese Mandarin sample, the results of linear regression in the study of Zhao et al. (2022) showed that the MFCC7 parameter predicted the PHQ-9 score, and the MFCC9 parameter predicted the HAMD anxiety score. The results of the multiple linear regression analysis by Silva et al. (2024) indicate that, among the parameters analyzed (mean, mode, and standard deviation of F0, Jitter, Shimmer, glottal to noise excitation ratio, smoothed cepstral peak prominence, and spectral tilt), the Jitter parameter and smoothed cepstral peak prominence serve as predictors of depression measured by the BDI-II. In one longitudinal studies (Wadle et al., 2024), voice characteristics were monitored over a three-week period in participants undergoing sleep deprivation therapy. Results from multilevel linear regression analysis indicated that speech pauses and pitch variability were significant predictors of depression severity (MADRS), whereas speech rate was not a significant predictor. Different types of speech tasks (reading, sustained vowel phonation and continuous speech) within the same vocal analysis, analyzed parameters and depression rating scales may underlie discrepancies in results.

Some authors use a different prediction paradigm, as traditional regression analysis predicts a functional relationship between voice and speech characteristics and depression scores (Cummins et al., 2020). Shin et al. (2021) show that a multilayer processing method, as a machine-learning approach, provides the best recognition results with an accuracy of 65.6%. Also, as one of the scarce studies that includes participants with minor depression, the mentioned research found that this method can differentiate between participants with minor and major depressive disorder. Du et al. (2022) analyzed acoustic voice characteristics (voice quality, prosodic and spectral features) based on reading a text in a smaller sample of participants with depression. Principal component analysis was first applied, followed by a multilayer perceptron to establish and compare a classification model with traditional classifiers. The multilayer perceptron provided the best results with an accuracy of 0.875. In addition to traditional machine learning, there have recently been attempts to detect depression using neural networks. A longitudinal study (Wang et al., 2023) shows that a neural network-based model can predict depression severity based on acoustic voice characteristics, with a correlation coefficient of 0.684. A model based on a convolutional neural network (Chlasta et al., 2019) also showed the ability to recognize depression from speech with an accuracy of 77%. The variability in prediction accuracy across these models can be attributed to differences in sample characteristics, analysis methods, and selected features, and speech tasks.

Research Hypothesis

Although traditional machine learning and neural network models often show a high degree of accuracy in the prediction of depression based on acoustic characteristics, some authors point out that simpler models (like logistic or linear regression) are efficient enough and do not differ significantly from more complex models (like neural networks), provided the data are clear and aligned with the characteristics of what is being examined (Rudin, 2019). We share the view that linear regression retains an important role and enables a clear interpretation of the relationship between predictors and criteria, providing insight into which voice characteristics contribute most to the severity of depression. Our hypothesis is that voice characteristics (both perceptual and acoustic) will have a predictive value in determining the severity of depression. If voice characteristics are found to have potential as an objective biomarker of depression in our sample through regression analysis, this could contribute to the creation of artificial intelligence AI models in the future, allowing for comparison and deepening of this knowledge.

Aim of the Research

Our research aims to determine whether specific voice characteristics, perceptual and acoustic, can predict the severity of depression measured by the MADRS scale in a sample of Serbian speakers.

Materials and Methods

The sample

The study included 100 participants, with the experimental group consisting of 45 participants diagnosed with a depressive disorder and the control group consisting of 55 participants without a depressive disorder. The experimental group included three subgroups based on depression severity: mild, moderate, and severe. Each subgroup consisted of 15 participants. The sample included only participants aged between 18 and 64 years, with no comorbid psychiatric disorders or somatic diseases (which could affect the voice) and professional voice users with fewer than ten years of work experience. The participants were native speakers of Serbian. Since physiological changes associated with aging can affect the vocal cords and voice quality (Petrović-Lazić and Ilić Savić, 2023; Petrović-Lazić et al., 2008), elderly participants were not selected. A psychiatrist made the diagnosis based on an interview and the guidelines provided in the DSM-V (APA, 2013) and additionally applied the MADRS scale to determine the severity of depression. The experimental and control groups were not statistically significantly different in gender ($\chi 2 = 0.756$, $p > 0.05$) or age ($F = 0.080$, $p > 0.05$).

Table 1. *Sample characteristics*

Variable		Experimental group	Control group
Number of participants	N = 100	45	55
Gender	Male	15	23
	Female	30	32
Age (M ± SD)		45.82 ± 12.520	41.29 ± 12.060
Smoking status	Yes	28	14
	No	17	41
Depression severity	Without depression symptoms	0	55
	Mild depression	15	0
	Moderate depression	15	0
	Severe depression	15	0

The participants in the experimental group were selected based on the psychiatrist's recommendation following the diagnostic and research criteria. The participants in the control group were conveniently selected from Kragujevac and its surroundings, matched by gender and age with the experimental group participants. Data on diagnosis, the absence of comorbid psychiatric and somatic conditions, and sociodemo-

graphic data (gender, age, profession, smoking status) were obtained from medical records and interviews.

Procedure and instruments

The study was approved by the Ethics Committee of the University Clinical Center Kragujevac (no. 01/21-422) and was conducted at the Psychiatry Clinic between 2021 and 2023. The research conducted with each participant individually started only after they received a detailed explanation of the purpose and procedure of the study and signed the informed consent for participation in the research.

The recording was done in a room isolated from distractions and noise. A speech therapist conducted the voice recording, while a psychiatrist administered the MADRS scale to obtain data on depression severity.

The severity of depression was assessed using the Montgomery-Asberg Depression Rating Scale (MADRS; Montgomery and Asberg, 1979). The scale was validated for the Serbian-speaking population (Mihajlović et al., 2021) and showed high internal reliability (α = 0.84). It includes ten items with a seven-point Likert-type scale (0 - no symptoms; 6 - severely expressed symptoms). The items primarily assess the main symptoms of depression (sadness, tension, concentration, fatigue, loss of interest, pessimistic thoughts), as well as somatic symptoms (appetite, sleep). The psychiatrist rates one item, while the participants self-assess the remaining nine. In our study, the Cronbach's alpha value indicates that the scale is highly reliable (α = 0.97).

The Multidimensional Voice Program (MDVP) by Kay Elemetrics, model 4300, was used to analyze acoustic voice characteristics. This software allows for the acoustic analysis of 33 parameters in numerical and graphical form (Petrović-Lazić, 2021). The participants had a task to sustain the vowel /a/ for approximately three seconds. A Sony ECM-T150 microphone was used for recording, positioned about 5 cm from the participant's mouth.

We analyzed 15 acoustic parameters in the domains of frequency variability (F0, Fhi, Flo, vF0, PFR, STD, Jitt, PPQ), intensity variability (ShdB, Shim, vAm, APQ), and noise and tremor estimation (NHR, VTI, SPI). The voice characteristics were chosen based on their frequent use in examining voice acoustics in depression and, generally, in voice pathology.

Perceptual voice characteristics were analyzed using the GRBAS scale (Isshiki et al., 1969). The participants had a task to read a phonetically balanced text. Each parameter of the GRBAS scale – G (grade) for overall hoarseness, R (roughness) for vocal roughness, B (breathiness) for vocal breathiness, A (asthenia) for vocal weakness, and S (strain) for vocal tension – was independently assessed by three voice pathologists using a four-point rating scale (0 = normal; 1 = mild/low degree; 2 = moderate/moderate degree; 3 = severe/high degree), after which the average score was calculated.

Table 2. *Kappa coefficients of inter-rater agreement for perceptual characteristics of voice between pairs of raters*

		1 vs 2	1 vs 3	2 vs 3
G	Kappa	0.835	0.888	0.831
	p	**0.000**	**0.000**	**0.000**
R	Kappa	0.615	0.639	0.852
	p	**0.000**	**0.000**	**0.000**
B	Kappa	0.708	0.752	0.693
	p	**0.000**	**0.000**	**0.000**
A	Kappa	0.785	0.803	0.718
	p	**0.000**	**0.000**	**0.000**
S	Kappa	0.680	0.747	0.848
	p	**0.000**	**0.000**	**0.000**

Kappa values indicated substantial agreement between raters across perceptual voice characteristics, with the strongest agreement for parameter G (almost perfect). All values were statistically significant (p = 0.000).

While some researchers argue that sustained vowel phonation is a more precise measure for objective voice analysis (Gerratt et al., 2016; Nguyen et al., 2024), others suggest that continuous speech is more suitable for the perceptual identification of hoarseness due to the greater number of vocal fold vibrations and increased vocal strain (Stráník, 2014), which justifies our choice of speech tasks within both vocal analyses.

The effectiveness of the MDVP software and the GRBAS scale for assessing voice quality was confirmed by research conducted in the Serbian-speaking area (e.g. Arsenić et al., 2021; Calić et al., 2022b; Petrović-Lazić et al., 2016; Šehović et al., 2017).

Table 3. *Analyzed voice characteristics*

Domains of voice characteristics	Voice characteristics labels	Explanation of labels
Parameters of frequency variability	F0	average fundamental frequency
	Fhi	highest fundamental frequency
	Flo	lowest fundamental frequency
	vF0	coefficient of fundamental frequency variation
	PFR	phonatory fundamental frequency range
	STD	standard deviation of the fundamental frequency
	Jitt	Jitter percent
	PPQ	pitch perturbation quotient
Parameters of intensity variability	ShdB	Shimmer in dB
	Shim	Shimmer percent
	vAm	peak amplitude variation
	APQ	amplitude perturbation quotient
Parameters of noise and tremor estimation	NHR	noise-to-harmonic ratio
	VTI	voice turbulence index
	SPI	soft phonation index
Perceptual parameters	G	overall grade of hoarseness
	R	roughness in voice
	B	breathiness in voice
	A	asthenia in voice
	S	strain in voice

Statistical data analysis

The analyses included both descriptive and analytical statistical measures. The following descriptive measures were presented: minimum, maximum, arithmetic mean, standard deviation, median, and interquartile range. Based on the results of the Kolmogorov-Smirnov test, which indicated that the distribution of the obtained measures significantly deviated from a normal distribution, nonparametric statistical methods were used. The Kruskal-Wallis test was applied to examine differences in numerical variables between groups, while Dunn-Bonferroni post hoc analyses were used to determine differences between specific subgroup pairs. MANCOVA was additionally performed to assess whether subgroups of different depression severity levels differed in voice characteristics, after adjusting for the effects of gender, age and smoking status as covariates. Hierarchical regression analysis was conducted to assess the predictive role of independent variables on the dependent variable. The level of statistical significance was set at $p \leq 0.05$.

The statistical analysis was performed using the Statistical Package for the Social Sciences (SPSS), version 26 (2019).

Results

Descriptive measures and testing differences in voice characteristics

Table 4 presents the descriptive measures for acoustic voice characteristics in participants with different levels of depression symptoms and determines the significance of differences among them.

Table 4. *Descriptive measures for acoustic voice characteristics in participants with different levels of depression symptoms and testing differences*

	Groups	N	Min	Max	M(SD)	95% CI	Mdn (IQR)	Kruskal-Wallis test
Fo	none	55	84.962	269.600	171.356(51.017)	157.564-185.148	178.709(83.514)	KW = 5.163 df= 3 p = 0.160
	mild	15	100.311	210.129	161.282(38.841)	139.772-182.791	175.200(68.871)	
	moderate	15	107.405	213.997	146.698(36.774)	126.334-167.063	145.510(57.930)	
	severe	15	102.352	214.800	144.945(36.488)	124.739-165.151	137.519(45.557)	
Fhi	none	55	99.644	314.954	186.281(58.509)	170.464-202.099	195.111(97.662)	KW = 0.800 df = 3 p = 0.849
	mild	15	110.860	238.701	181.149(42.640)	157.536-204.763	193.614(82.295)	
	moderate	15	116.727	245.661	179.665(46.745)	153.779-205.552	183.276(95.493)	
	severe	15	111.579	253.318	174.079(46.312)	148.433-199.726	172.195(71.826)	
Flo	none	55	77.257	251.699	155.194(46.442)	142.639-167.749	162.680(73.111)	KW =11.408 df = 3 **p = 0.010**
	mild	15	87.722	190.773	142.106(36.012)	122.163-162.049	158.641(69.132)	
	moderate	15	74.000	180.463	125.373(33.252)	106.958-143.787	111.216(59.861)	
	severe	15	68.657	200.858	118.990(35.824)	99.151-138.828	112.147(43.755)	
STD	none	55	.893	13.607	2.951(2.038)	2.400-3.502	2.484(1.769)	KW = 21.245 df = 3 **p = 0.000**
	mild	15	1.498	6.660	4.227(1.494)	3.400-5.054	3.952(2.563)	
	moderate	15	1.189	16.251	5.969(3.887)	3.817-8.122	6.143(4.521)	
	severe	15	1.528	40.985	9.972(11.951)	3.354-16.590	5.528(5.350)	
PFR	none	55	1.000	11.000	4.091(2.263)	3.479-4.703	3.000(4.000)	KW = 21.402 df = 3 **p = 0.000**
	mild	15	3.000	10.000	5.133(2.200)	3.915-6.351	4.000(4.000)	
	moderate	15	3.000	16.000	7.800(4.491)	5.313-10.287	7.000(9.000)	
	severe	15	3.000	18.000	8.533(4.984)	5.773-11.293	7.000(6.000)	
vF0	none	55	.636	6.426	1.723(0.933)	1.470-1.975	1.513(1.009)	KW = 32.671 df = 3 **p = 0.000**
	mild	15	1.207	5.380	2.718(1.095)	2.112-3.325	2.710(1.746)	
	moderate	15	1.060	15.130	4.199(3.428)	2.300-6.097	3.536(2.010)	
	severe	15	1.449	25.212	6.167(6.930)	2.329-10.004	3.673(3.166)	
Jitt	none	55	.266	1.931	0.626(0.346)	0.533-0.720	0.557(0.346)	KW = 39.779 df = 3 **p = 0.000**
	mild	15	0.389	3.777	1.400(0.889)	0.907-1.892	1.391(0.982)	
	moderate	15	0.535	5.172	1.573(1.257)	0.877-2.269	1.165(0.862)	
	severe	15	0.373	4.223	2.210(1.310)	1.484-2.935	2.012(2.259)	
ShdB	none	55	0.106	0.897	0.313(0.159)	0.270-0.356	0.262(0.163)	KW = 36.809 df = 23 **p = 0.000**
	mild	15	0.220	0.951	0.449(0.199)	0.339-0.559	0.370(0.205)	
	moderate	15	0.286	1.395	0.632(0.315)	0.458-0.807	0.509(0.236)	
	severe	15	0.266	1.144	0.616(0.236)	0.485-0.746	0.597(0.362)	
Shim	none	55	1.225	9.720	3.488(1.753)	3.015-3.962	2.933(1.879)	KW = 36.971 df = 3 **p = 0.000**
	mild	15	2.500	10.650	5.042(2.120)	3.868-6.216	4.288(2.386)	
	moderate	15	3.303	12.408	6.662(2.902)	5.054-8.269	5.449(2.745)	
	severe	15	2.963	11.369	6.735(2.398)	5.407-8.062	6.795(4.150)	

	Groups	N	Min	Max	M(SD)	95% CI	Mdn (IQR)	Kruskal-Wallis test
APQ	none	55	.878	6.803	2.724(1.224)	2.393-3.054	2.437(1.298)	KW = 39.012 df = 3 **p = 0.000**
	mild	15	2.001	7.207	3.762(1.429)	2.970-4.553	3.614(1.424)	
	moderate	15	2,754	9.337	4.951(1.925)	3.885-6.017	4.376(1.651)	
	severe	15	2.924	7.744	4.915(1.461)	4.106-5.724	5.006(2.875)	
PPQ	none	55	0.150	1.112	0.354(0.191)	0.303-0.406	0.294(0.211)	KW = 41.397 df = 3 **p = 0.000**
	mild	15	0.240	2.506	0.834(0.590)	0.508-1.161	0.739(0.581)	
	moderate	15	0.303	3.164	0.904(0.734)	0.498-1.311	0.673(0.552)	
	severe	15	0.214	2.539	1.328(0.817)	0.876-1.781	1.188(1.553)	
vAm	none	55	4.377	26.488	10.191(5.328)	8.750-11.631	8.629(5.410)	KW = 41.945 df = 3 **p = 0.000**
	mild	15	6.127	36.293	19.878(8.724)	15.046-24.709	18.997(15.316)	
	moderate	15	7.717	43.602	21.397(8.825)	16.510-26.284	21.026(7.901)	
	severe	15	11.915	42.912	21.160(8.523)	16.441-25.880	19.176(13.713)	
NHR	none	55	0.106	0.250	0.136(0.023)	0.130-0.143	0.136(0.025)	KW = 23.727 df = 3 **p = 0.000**
	mild	15	0.114	0.199	0.150(0.025)	0.137-0.164	0.143(0.022)	
	moderate	15	0.124	0.274	0.176(0.038)	0.155-0.197	0.165(0.033)	
	severe	15	0.118	0.270	0.173(0.050)	0.146-0.201	0.155(0.077)	
VTI	none	55	0.014	0.106	0.055(0.017)	0.051-0.060	0.054(0.025)	KW = 1.987 df = 3 p = 0.575
	mild	15	0.024	0.088	0.056(0.017)	0.047-0.066	0.055(0.024)	
	moderate	15	0.026	0.095	0.059(0.019)	0.048-0.069	0.060(0.032)	
	severe	15	0.021	0.108	0.063(0.022)	0.051-0.075	0.067(0.034)	
SPI	none	55	1.697	32.791	6.593(4.481)	5.381-7.804	6.006(3.365)	KW = 19.451 df = 3 **p = 0.000**
	mild	15	2.882	18.861	9.162(5.184)	6.291-12.033	7.473(10.353)	
	moderate	15	4.331	19.894	9.206(4.553)	6.685-11.727	8.305(6.033)	
	severe	15	4.456	16.019	10.773(3.413)	8.883-12.663	10.242(6.383)	

Notes: N = number of participants; Min = minimum; Max = maximum; M = arithmetic mean; SD = standard deviation; 95% CI = 95% confidence interval (lower and upper bound); Mdn = median; IQR = interquartile range; KW = Kruskal-Wallis test; df = degrees of freedom; p = statistical significance

The results of the Kruskal-Wallis test indicate statistically significant differences among participants with different levels of depression symptoms (none, mild, moderate, severe) for all analyzed acoustic voice characteristics ($p \leq 0.01$) except for the average fundamental frequency (F0), the highest fundamental frequency (Fhi), and the voice turbulence index (VTI) ($p > 0.05$).

Dunn-Bonferroni analyses were applied to more precisely determine which pairs of subgroups, according to depression severity, show differences in acoustic voice characteristics (Table 5).

Table 5. *Results of the Kruskal-Wallis test with Dunn-Bonferroni analyses examining the differences between pairs of subgroups according to depression severity with regard to acoustic voice characteristics*

			Test statistic	Std.Error	Std. Test Statistic	p	Adj. p
Flo	severe	moderate	3.533	10.593	0.334	0.739	1.000
	severe	mild	15.733	10.593	1.485	0.137	0.825
	severe	none	23.897	8.451	2.828	**0.005**	0.028
	moderate	mild	12.200	10.593	1.152	0.249	1.000
	moderate	none	20.364	8.451	2.410	**0.016**	0.096
	mild	none	8.164	8.451	0.966	0.334	1.000
STD	none	mild	-21.970	8.451	-2.600	**0.009**	0.056
	none	moderate	-28.636	8.451	-3.389	**0.001**	0.004
	none	severe	-28.970	8.451	-3.428	**0.001**	0.004
	mild	moderate	-6.667	10.593	-0.629	0.529	1.000
	mild	severe	-7.000	10.593	-0.661	0.509	1.000
	moderate	severe	-.333	10.593	-0.031	0.975	1.000
PFR	none	mild	-14.721	8.359	-1.761	0.078	0.469
	none	moderate	-27.688	8.359	-3.312	**0.001**	0.006
	none	severe	-31.955	8.359	-3.823	**0.000**	0.001
	mild	moderate	-12.967	10.478	-1.237	0.216	1.000
	mild	severe	-17.233	10.478	-1.645	0.100	0.600
	moderate	severe	-4.267	10.478	-0.407	0.684	1.000
vF0	none	mild	-26.018	8.451	-3.079	**0.002**	0.012
	none	moderate	-35.352	8.451	-4.183	**0.000**	0.000
	none	severe	-36.752	8.451	-4.349	**0.000**	0.000
	mild	moderate	-9.333	10.593	-0.881	0.378	1.000
	mild	severe	-10.733	10.593	-1.013	0.311	1.000
	moderate	severe	-1.400	10.593	-0.132	0.895	1.000
Jitt	none	mild	-30.855	8.451	-3.651	**0.000**	0.002
	none	moderate	-34.421	8.451	-4.073	**0.000**	0.000
	none	severe	-43.088	8.451	-5.099	**0.000**	0.000
	mild	moderate	-3.567	10.593	-0.337	0.736	1.000
	mild	severe	-12.233	10.593	-1.155	0.248	1.000
	moderate	severe	-8.667	10.593	-0.818	0.413	1.000
ShdB	none	mild	-22.539	8.451	-2.667	**0.008**	0.046
	none	moderate	-38.273	8.451	-4.529	**0.000**	0.000
	none	severe	-40.339	8.451	-4.774	**0.000**	0.000
	mild	moderate	-15.733	10.593	-1.485	0.137	0.825
	mild	severe	-17.800	10.593	-1.680	0.093	0.557
	moderate	severe	-2.067	10.593	-0.195	0.845	1.000
Shim	none	mild	-23.533	8.451	-2.785	**0.005**	0.032
	none	moderate	-38.067	8.451	-4.505	**0.000**	0.000
	none	severe	-40.400	8.451	-4.781	**0.000**	0.000
	mild	moderate	-14.533	10.593	-1.372	0.170	1.000
	mild	severe	-16.867	10.593	-1.592	0.111	0.668
	moderate	severe	-2.333	10.593	-0.220	0.826	1.000

			Test statistic	Std.Error	Std. Test Statistic	p	Adj. p
APQ	none	mild	-22.294	8.451	-2.638	**0.008**	0.050
	none	moderate	-39.961	8.451	-4.729	**0.000**	0.000
	none	severe	-41.261	8.451	-4.883	**0.000**	0.000
	mild	moderate	-17.667	10.593	-1.668	0.095	0.572
	mild	severe	-18.967	10.593	-1.790	0.073	0.440
	moderate	severe	-1.300	10.593	-0.123	0.902	1.000
PPQ	none	mild	-31.091	8.451	-3.679	**0.000**	0.001
	none	moderate	-35.891	8.451	-4.247	**0.000**	0.000
	none	severe	-43.624	8.451	-5.162	**0.000**	0.000
	mild	moderate	-4.800	10.593	-0.453	0.650	1.000
	mild	severe	-12.533	10.593	-1.183	0.237	1.000
	moderate	severe	-7.733	10.593	-0.730	0.465	1.000
vAm	none	mild	-34.370	8.451	-4.067	**0.000**	0.000
	none	severe	-38.836	8.451	-4.596	**0.000**	0.000
	none	moderate	-39.703	8.451	-4.698	**0.000**	0.000
	mild	severe	-4.467	10.593	-0.422	0.673	1.000
	mild	moderate	-5.333	10.593	-0.503	0.615	1.000
	severe	moderate	.867	10.593	0.082	0.935	1.000
NHR	none	mild	-17.476	8.448	-2.069	**0.039**	0.232
	none	severe	-25.776	8.448	-3.051	**0.002**	0.014
	none	moderate	-36.142	8.448	-4.278	**0.000**	0.000
	mild	severe	-8.300	10.590	-0.784	0.433	1.000
	mild	moderate	-18.667	10.590	-1.763	0.078	0.468
	severe	moderate	10.367	10.590	0.979	0.328	1.000
SPI	none	mild	-17.485	8.451	-2.069	**0.039**	0.231
	none	moderate	-20.085	8.451	-2.377	**0.017**	0.105
	none	severe	-33.885	8.451	-4.010	**0.000**	0.000
	mild	moderate	-2.600	10.593	-0.245	0.806	1.000
	mild	severe	-16.400	10.593	-1.548	0.122	0.730
	moderate	severe	-13.800	10.593	-1.303	0.193	1.000

Notes: p = statistical significance; Adj. p = adjusted statistical significance

The results indicate significant differences between participants without depression and those with depression (mild, moderate, severe) for all acoustic voice characteristics ($p < 0.05$) except for the lowest fundamental frequency (Flo) and the fundamental frequency range (PFR) ($p > 0.05$) between participants without depression and those with mild depression, while no significant differences ($p > 0.05$) were observed between subgroups of participants with mild, moderate, and severe depression.

Table 6. *Descriptive measures for perceptual voice characteristics in participants with different levels of depression symptoms and testing differences*

	Groups	N	Min	Max	M(SD)	95% CI	Mdn (IQR)	Kruskal-Wallis test
G	none	55	0.000	1.000	0.055(0.229)	-0.007- 0.117	0.000(0.000)	KW = 32.731 df = 3 **p = 0.000**
	mild	15	0.000	1.000	0.156(0.330)	-0.027- 0.338	0.000(0.000)	
	moderate	15	0.000	1.667	0.511(0.589)	0.185- 0.837	0.000(1.000)	
	severe	15	0.000	2.333	0.911(0.791)	0.473- 1.349	1.000(1.667)	
R	none	55	0.000	1.000	0.103(0.300)	0.022- 0.184	0.000(0.000)	KW = 22.003 df = 3 **p = 0.000**
	mild	15	0.000	0.667	0.067(0.187)	-0.037- 0.170	0.000(0.000)	
	moderate	15	0.000	1.667	0.267(0.507)	-0.014- 0.547	0.000(0.667)	
	severe	15	0.000	2.000	0.667(0.678)	0.291- 1.042	0.667(1.000)	
B	none	55	0.000	1.000	0.097(0.246)	0.031- 0.163	0.000(0.000)	KW = 39.004 df = 3 **p = 0.000**
	mild	15	0.000	1.667	0.533(0.615)	0.193- 0.874	0.333(1.000)	
	moderate	15	0.000	1.333	0.378(0.517)	0.091- 0.664	0.000(1.000)	
	severe	15	0.000	2.000	1.111(0.600)	0.779- 1.443	1.000(1.000)	
A	none	55	0.000	1.000	0.079(0.248)	0.012-0.146	0.000(0.000)	KW = 33.526 df = 3 **p = 0.000**
	mild	15	0.000	1.000	0.267(0.402)	0.044- 0.489	0.000(0.667)	
	moderate	15	0.000	1.333	0.511(0.486)	0.242- 0.780	0.667(1.000)	
	severe	15	0.000	2.000	0.889(0.626)	0.542- 1.235	1.000(1.333)	
S	none	55	0.000	1.000	0.091(0.276)	0.016- 0.165	0.000(0.000)	KW = 30.947 df = 3 **p = 0.000**
	mild	15	0.000	1.000	0.244(0.320)	0.067- 0.422	0.000(0.333)	
	moderate	15	0.000	1.000	0.467(0.433)	0.227- 0.706	0.667(1.000)	
	severe	15	0.000	1.000	0.711(0.452)	0.461- 0.961	1.000(1.000)	

Notes: N = number of participants; Min = minimum; Max = maximum; M = arithmetic mean; SD = standard deviation; 95% CI = 95% confidence interval (lower and upper bound); Mdn = median; IQR = interquartile range; KW = Kruskal-Wallis test; df = degrees of freedom; p = statistical significance

The results of the Kruskal-Wallis test show statistically significant differences between participants with different levels of depression symptoms for all analyzed perceptual voice characteristics ($p < 0.001$). In addition, Dunn-Bonferroni analyses were conducted to more precisely determine which pairs of subgroups, according to depression severity, show differences in perceptual voice characteristics (Table 7).

Table 7. *Results of the Kruskal-Wallis test with Dunn-Bonferroni analyses examining the differences between pairs of subgroups according to the severity of depression with regard to perceptual voice characteristics*

			Test statistic	Std.Error	Std. Test Statistic	p	Adj. p
G	none	mild	-5.921	6.214	-0.953	0.341	1.000
	none	moderate	-20.621	6.214	-3.318	**0.001**	0.005
	none	severe	-32.488	6.214	-5.228	**0.000**	0.000
	mild	moderate	-14.700	7.790	-1.887	0.059	0.355
	mild	severe	-26.567	7.790	-3.410	**0.001**	0.004
	moderate	severe	-11.867	7.790	-1.523	0.128	0.766
R	none	mild	-.064	6.119	-0.010	0.992	1.000
	none	moderate	-7.797	6.119	-1.274	0.203	1.000
	none	severe	-27.897	6.119	-4.559	**0.000**	0.000
	mild	moderate	-7.733	7.670	-1.008	0.313	1.000
	mild	severe	-27.833	7.670	-3.629	**0.000**	0.002
	moderate	severe	-20.100	7.670	-2.621	**0.009**	0.053

			Test statistic	Std.Error	Std. Test Statistic	p	Adj. p
B	none	moderate	-13.803	7.299	-1.891	0.059	0.352
	none	mild	-20.303	7.299	-2.782	**0.005**	0.032
	none	severe	-44.136	7.299	-6.047	**0.000**	0.000
	moderate	mild	6.500	9.149	0.710	0.477	1.000
	moderate	severe	-30.333	9.149	-3.315	**0.001**	0.005
	mild	severe	-23.833	9.149	-2.605	**0.009**	0.055
A	none	mild	-10.282	6.902	-1.490	0.136	0.818
	none	moderate	-23.815	6.902	-3.450	**0.001**	0.003
	none	severe	-36.448	6.902	-5.281	**0.000**	0.000
	mild	moderate	-13.533	8.652	-1.564	0.118	0.707
	mild	severe	-26.167	8.652	-3.024	**0.002**	0.015
	moderate	severe	-12.633	8.652	-1.460	0.144	0.866
S	none	mild	-13.733	7.028	-1.954	0.051	0.304
	none	moderate	-23.633	7.028	-3.363	**0.001**	0.005
	none	severe	-35.300	7.028	-5.022	**0.000**	0.000
	mild	moderate	-9.900	8.811	-1.124	0.261	1.000
	mild	severe	-21.567	8.811	-2.448	**0.014**	0.086
	moderate	severe	-11.667	8.811	-1.324	0.185	1.000

Notes: p = statistical significance; Adj. p = adjusted statistical significance

The results indicate statistically significant differences between participants without depression and participants with depression for all perceptual voice characteristics ($p < 0.01$) except for hoarseness (G), roughness (R), asthenia (A), and strain (S) ($p > 0.05$) between participants without depression and those with mild depression, as well as roughness (R) and breathiness (B) ($p > 0.05$) between participants without depression and those with moderate depression. Significant differences for all parameters ($p < 0.05$) were found between participants with mild and severe depression, while for R and B parameters ($p < 0.01$), differences were determined between participants with moderate and severe depression. There were no significant differences between participants with mild and moderate depression in any perceptual parameters ($p > 0.05$).

MANCOVA was performed to assess whether subgroups of different depression severity levels differed in voice characteristics, after adjusting for the effects of gender, age and smoking status as covariates (Table 8).

Table 8. *Multivariate effects of gender, age, smoking status and depression severity on acoustic and perceptual voice characteristics*

Acoustic characteristics	Wilks' Lambda	F	df1	df2	p	η^2
gender	0.257	15.25	15	79	**0.000**	0.743
age	0.686	2.41	15	79	**0.006**	0.314
smoking status	0.850	0.93	15	79	0.539	0.150
depression severity	0.338	2.31	45	235.5	**0.000**	0.303
Perceptual characteristics						
gender	0.991	0.163	5	89	0.976	0.009
age	0.813	4.102	5	89	**0.002**	0.187
smoking status	0.935	1.235	5	89	0.300	0.065
depression severity	0.329	8.123	15	246.1	**0.000**	0.309

Notes: df1, df2 = degrees of freedom; p = statistical significance; η^2 = Partial Eta Squared

The results of MANCOVA test show that gender has a statistically significant effect on the overall acoustic characteristics of the voice ($p < 0.001$) with a very large effect size ($\eta^2 = 0.743$). Age also has a

statistically significant, but moderate effect ($p < 0.01$; $\eta^2 = 0.314$), while smoking status has no statistically significant effect ($p > 0.05$). Depression severity has a statistically significant effect ($p < 0.001$), with a moderate effect size ($\eta^2 = 0.303$).

Regarding the perceptual voice characteristics, the MANCOVA test indicates that gender and smoking status have no statistically significant effect ($p > 0.05$). Age has a statistically significant but small effect ($p < 0.01$; $\eta^2 = 0.187$), while depression severity shows a statistically significant effect ($p < 0.001$), with a moderate effect size ($\eta^2 = 0.309$).

Table 9. *Univariate effects of gender, age, smoking status and depression severity on acoustic and perceptual voice characteristics*

Voice characteristics	gender				age				smoking status				depression severity			
	F	df	p	η²	F	df	p	η²	F	df	p	η²	F	df	p	η²
F0	177.040	1	**0.000**	0.656	2.180	1	0.143	0.023	4.316	1	**0.041**	0.044	4.326	3	**0.007**	0.122
Fhi	140.729	1	**0.000**	0.602	0.331	1	0.566	0.004	3.398	1	0.068	0.035	0.859	3	0.465	0.027
Flo	90.433	1	**0.000**	0.493	2.268	1	0.135	0.024	5.815	1	**0.018**	0.059	4.982	3	**0.003**	0.138
STD	10.207	1	**0.002**	0.099	2.848	1	0.095	0.030	1.229	1	0.271	0.013	7.402	3	**0.000**	0.193
PFR	6.952	1	**0.010**	0.070	0.290	1	0.592	0.003	1.062	1	0.306	0.011	9.473	3	**0.000**	0.234
vFo	2.402	1	0.125	0.025	4.337	1	**0.040**	0.045	3.194	1	0.077	0.033	7.145	3	**0.000**	0.187
Jitt	0.802	1	0.373	0.009	3.179	1	0.078	0.033	0.063	1	0.803	0.001	14.383	3	**0.000**	0.317
ShdB	0.433	1	0.512	0.005	1.444	1	0.233	0.015	0.040	1	0.842	0.000	11.937	3	**0.000**	0.278
Shim	1.129	1	0.291	0.012	1.684	1	0.198	0.018	0.010	1	0.920	0.000	12.288	3	**0.000**	0.284
APQ	4.461	1	0.037	0.046	2.869	1	0.094	0.030	0.268	1	0.606	0.003	12.986	3	**0.000**	0.295
PPQ	0.577	1	0.450	0.006	3.654	1	0.059	0.038	0.001	1	0.978	0.000	13.912	3	**0.000**	0.310
vAm	3.113	1	0.081	0.032	1.944	1	0.167	0.020	0.026	1	0.873	0.000	14.374	3	**0.000**	0.317
NHR	0.164	1	0.686	0.002	0.363	1	0.548	0.004	0.363	1	0.548	0.004	7.704	3	**0.000**	0.199
VTI	0.485	1	0.488	0.005	3.742	1	0.056	0.039	2.916	1	0.091	0.030	0.929	3	0.430	0.029
SPI	3.726	1	0.057	0.039	17.573	1	**0.000**	0.159	0.204	1	0.653	0.002	3.214	3	**0.026**	0.094
G	0.301	1	0.584	0.003	8.602	1	**0.004**	0.085	0.553	1	0.459	0.006	14.871	3	**0.000**	0.324
R	0.006	1	0.939	0.000	3.690	1	0.058	0.038	1.014	1	0.317	0.011	7.088	3	**0.000**	0.186
B	0.154	1	0.696	0.002	6.202	1	**0.015**	0.063	0.021	1	0.884	0.000	18.283	3	**0.000**	0.371
A	0.060	1	0.807	0.001	8.478	1	**0.005**	0.084	1.734	1	0.191	0.018	15.032	3	**0.000**	0.327
S	0.230	1	0.632	0.002	0.316	1	0.575	0.003	3.103	1	0.081	0.032	16.110	3	**0.000**	0.342

Notes: df = degrees of freedom; p = statistical significance; η² = Partial Eta Squared

The effect of gender was statistically significant for the following acoustic characteristics: F0 ($p < 0.001$, $\eta^2 = 0.656$), Fhi ($p < 0.001$, $\eta^2 = 0.602$), Flo ($p < 0.001$, $\eta^2 = 0.493$), STD ($p < 0.01$, $\eta^2 = 0.099$) and PFR ($p = 0.01$, $\eta^2 = 0.070$), while no statistically significant effects were found for any of the perceptual characteristics ($p > 0.05$). Age had a significant effect on vF0 ($p < 0.05$, $\eta^2 = 0.045$) and SPI ($p < 0.001$, $\eta^2 = 0.159$) among the acoustic characteristics, and on G ($p < 0.01$, $\eta^2 = 0.085$), B ($p < 0.05$, $\eta^2 = 0.063$) and A ($p < 0.01$, $\eta^2 = 0.084$) among the perceptual ones. Smoking status showed a significant effect on F0 ($p < 0.05$, $\eta^2 = 0.044$) and Flo ($p < 0.05$, $\eta^2 = 0.059$) but no significant effects on perceptual characteristics ($p > 0.05$). Regarding depression severity, statistically significant effects were observed for nearly all acoustic parameters ($p < 0.05$), except Fhi and VTI ($p > 0.05$), as well as for all perceptual parameters ($p < 0.001$), after controlling for gender, age, and smoking status.

Predictors of depression severity

A hierarchical regression analysis was used to determine the contribution of acoustic voice characteristics in predicting depression severity (Table 10).

Table 10. *Results of hierarchical regression analysis for predicting depression severity (MADRS score) based on acoustic voice characteristics*

Block		β	t	p	R	R^2	F (4/44)	P
1	gender	0.037	0.377	0.707	0.359	0.129	4.730(3/96)	**0.004**
	age	0.187	1.921	0.058				
	smoking status	-0.323	-3.368	**0.001**				
2	gender	0.147	1.021	0.310	0.750	0.563	5.371(15/81)	**0.000**
	age	-0.034	-0.368	0.714				
	smoking status	-0.127	-1.516	0.133				
	F0	-0.658	-1.524	0.131				
	Fhi	-0.271	-0.709	0.480				
	Flo	0.759	1.874	0.065				
	STD	0.067	0.179	0.858				
	PFR	0.519	1.897	0.061				
	vF0	0.016	0.042	0.966				
	Jitt	1.471	1.897	0.061				
	ShdB	-1.334	-1.975	0.052				
	Shim	1.086	1.323	0.189				
	APQ	0.322	0.674	0.502				
	PPQ	-1.235	-1.619	0.109				
	vAm	0.241	2.089	**0.040**				
	NHR	-0.182	-1.095	0.277				
	VTI	0.115	1.341	0.184				
	SPI	0.108	1.151	0.253				

Dependent variable: MADRS score

The hierarchical regression analysis was conducted in two blocks. The first block included sociodemographic variables (gender, age, smoking status), while acoustic voice characteristics were added in the second block along with the sociodemographic variables.

The results show that smoking status was a significant predictor of depression severity in the first block ($\beta = -0.323$, $t = -3.368$, $p < 0.01$). When voice characteristics were added in the second block, none of the sociodemographic variables were significant. However, the peak amplitude variation (vAm) acoustic parameter was found to be a statistically significant predictor ($\beta = 0.241$, $t = 2.089$, $p < 0.05$) of depression severity.

The contribution of perceptual voice characteristics to predicting depression severity was also tested using the hierarchical regression analysis (Table 11).

Table 11. *Results of regression analysis for predicting depression severity (MADRS score) based on perceptual voice characteristics*

Block		β	t	p	R	R^2	F (4/44)	P
1	gender	0.037	0.377	0.707	0.359	0.129	4.730(3/96)	**0.004**
	age	0.187	1.921	0.058				
	smoking status	-0.323	-3.368	**0.001**				
2	gender	0.061	0.967	0.336	0.809	0.654	27.626(5/91)	**0.000**
	age	-0.101	-1.426	0.157				
	smoking status	-0.183	-2.832	**0.006**				
	G	0.292	3.446	**0.001**				
	R	0.025	0.314	0.755				
	B	0.216	2.621	**0.010**				
	A	0.229	2.741	**0.007**				
	S	0.302	4.237	**0.000**				

Dependent variable: MADRS score

Hierarchical regression analysis was conducted in two blocks. The first block included sociodemographic variables (gender, age, smoking status), and in the second block, perceptual voice characteristics were introduced alongside the sociodemographic variables.

The results show that smoking status was a significant predictor of depression severity in the first block ($p < 0.01$). In the second block, when voice characteristics were included, the smoking status variable ($\beta = -0.183$, $t = -2.832$, $p < 0.01$), as well as the perceptual parameters G ($\beta = 0.292$, $t = 3.446$, $p < 0.01$), B ($\beta = 0.216$, $t = 2.621$, $p = 0.01$), A ($\beta = 0.229$, $t = 2.741$, $p < 0.01$), and S ($\beta = 0.302$, $t = 4.237$, $p < 0.001$), were found to be significant predictors of depression severity.

Discussion

Scarce literature available in the Serbian-speaking area suggests that there are statistically significant differences between participants with depression and those in the control group regarding certain voice and speech characteristics, such as parameters of frequency variability, amplitude variability, noise and tremor (Calić et al., 2022a), average intensity values (Ćuk-Jovanović, 2003), utterance duration (Ćuk-Jovanović, 2002), as well as the discriminative role of intensity variability parameters (Calić et al., 2022a). In our study, we aimed to conduct a deeper analysis to explore whether specific voice characteristics (perceptual and acoustic) can predict depression severity (MADRS score) by applying hierarchical regression analysis, incorporating variables that might affect the voice (gender, age, smoking status) and which are described as potential confounders in the literature (Hashim et al., 2017; Wang et al., 2023).

The results of the Kruskal-Wallis test show statistically significant differences in all perceptual and nearly all acoustic voice characteristics, except for F0 and Fhi in the frequency variability domain and the VTI parameter in the noise and tremor assessment domain, between participants with different levels of depression symptoms. The study by Silva et al. (2024), which also employed a sustained vowel phonation task, similarly found that the average F0 parameter did not differ between groups, unlike the Shimmer and Jitter parameters. Since the vocal task involved sustained vowel phonation, pitch-related features such as F0 and Fhi may have been less sensitive to emotional variation, compared to tasks that include continuous speech or reading, where intonation and lexical accentuation are more pronounced. For example, Wang et al. (2019) found that F0 varied across different speech tasks, including answering questions, reading, picture description, and video watching. Additionally, these findings may be partly explained by a gender effect that could have masked the potential impact of depression on these pitch-related features. Given that the Serbian vocal system includes stable phonation with clearly articulated, unreduced vowels (Nikolić, 2016), the VTI parameter, which measures the turbulent component of the voice signal, might not show significant differences precisely because of phonetic stability and the nature of the vocal task. However, it is also possible that these specific acoustic parameters are not sufficiently sensitive markers for detecting depression-related vocal changes. A more precise post hoc analysis revealed significant differences between participants without depression and those with depression, as expected. However, surprisingly, there were no significant differences in acoustic voice characteristics between participants with mild, moderate, and severe depression, while significant differences were found in perceptual voice characteristics. Differences were observed between participants with mild and severe depression (all analyzed perceptual parameters) and between participants with moderate and severe depression (roughness and breathiness), but not between participants with mild and moderate depression. This potentially indicates that the voice of participants with different levels of depression severity conveys a subjectively different auditory impression, which is why it is also important to analyze the acoustic correlates. A recent study (Menne et al., 2024) showed that the Shimmer parameter had higher average values in participants with moderate depression compared to those with mild depression. However, the differences were not statistically significant, as in our study. One of the scarce studies (Shin et al., 2021) that included participants with minor depression found that only the standard deviation of the fundamental frequency (STD) parameter differed between participants with minor and major depressive disorder out of the 21 analyzed characteristics.

Additionally, gender was found to significantly influence frequency variability parameters (F0, Fhi, Flo, STD and PFR), which is consistent with known physiological differences in vocal fold size and tension between males and females (Abitbol et al., 1999). This biological influence may overshadow subtle emotional effects on pitch. Smoking status also showed a significant effect on F0 and Flo, while age appeared to influence acoustic parameters vF0 and SPI, as well as perceptual voice characteristics (G, B,

and A). These findings are in line with Songur et al. (2025), who reported that age, rather than gender, influences perceptual voice characteristics, and with previous studies indicating an effect of smoking on F0 parameters (Ayoub et al., 2019). Nevertheless, the MANCOVA analysis indicated that depression severity significantly affected most acoustic (except Fhi and VTI) and all perceptual voice characteristics, even after controlling for gender, age, and smoking status. While Kruskal–Wallis analysis did not show group differences in F0, MANCOVA revealed a significant effect of depression severity on this parameter after controlling for gender, age, and smoking. This suggests that the effect of depression on F0 may be masked by stronger demographic influences, particularly gender. These findings point to a potential independent effect of depression severity on voice characteristics, beyond the influence of demographic variables such as age, gender, and smoking status. In other words, even after statistically controlling for variables known to affect voice parameters, differences in both acoustic and perceptual voice characteristics remained significant across levels of depression severity. This suggests that changes in voice may not be solely attributable to demographic factors, but could also reflect underlying psychopathological processes associated with depression. However, caution is warranted when interpreting these findings, as the cross-sectional nature of the study limits causal inferences.

The hierarchical regression analysis showed that among acoustic voice characteristics, the peak amplitude variation (vAm) from the second block was a significant predictor of depression severity (MADRS score). Although the smoking status variable was significant in the first model, it was not significant in the second model after adding acoustic voice characteristics, nor were the gender and age variables. These results are inconsistent with the results of multiple linear regression obtained by Silva et al. (2024), indicating that the Jitter parameter and the smoothed cepstral peak prominence were the predictors of depression. A possible explanation for this difference is that they used the Beck Depression Inventory, which is more focused on cognitive symptoms (Ignjatović Ristić et al., 2012; Kiss and Jenei, 2020), and they also had a higher proportion of participants with severe depression in their sample. High variations in peak-to-peak amplitude are associated with hypofunctional phonation, characterized by loose adduction (Laukkanen and Sundberg, 2008). Loose and shorter vocal folds, associated with lower F0, reduce adduction and thereby increase the amplitude of vocal fold vibrations (Laukkanen and Sundberg, 2008). Previous study (Calić et al., 2022a) also found that the peak amplitude variation parameter (vAm) had the highest discriminative value for the group of participants with depression, along with the amplitude perturbation quotient parameter (APQ) from the same domain. In the present study, APQ did not prove to be a significant predictor. However, the Shimmer in dB parameter (ShdB), which is related to APQ, was close to statistical significance. In a study by Quatieri and Malyska (2012), Shimmer was also found to be associated with depression severity measured by the HAMD scale, while Jitter was not significantly correlated. Future studies on larger samples that include an equal number of participants with different levels of depression severity could provide more precise significance.

In the group of perceptual voice characteristics, the significant predictors of depression severity were the G (hoarseness), B (breathiness), A (asthenia), and S (strain) parameters. Sahu and Espy-Wilson (2016) suggest that the vocal quality in depression is characterized by breathiness and creakiness, based on higher values of Jitter and Shimmer parameters. Wang et al. (2019) emphasize that vocal quality in depression may be characterized by vocal weakness due to the association between fundamental frequency parameters and overall muscle tension. In the model that includes perceptual voice characteristics, unlike the acoustic ones, smoking status emerged as a significant predictor of depression severity, while gender and age were not significant predictors in either model.

The obtained results confirm the existing literature on the predictive role of acoustic voice characteristics for depression, but they also preliminarily strengthen it by emphasizing the significant role of perceptual voice characteristics. Our study has several limitations. The first refers to the sample size, which should be larger in future studies to validate the results. It is important to increase the number of participants within each subgroup, especially those with severe depression, to improve the reliability of the regression analysis and reduce the risk of Type II error. Also, the sample should be expanded to include participants from different regions, and stratified random sampling should be applied to control groups to improve the generalizability of the results. Another limitation of the study is the lack of uniformity of the sample with respect to smoking status, in addition to gender and age, which may affect the generalizability of the results. Studies suggest that the prevalence of smoking is approximately twice as high among individuals with depression compared to those without (Lasser et al., 2000; Stubbs et al., 2018). Therefore,

it is important to control for the influence of smoking status in future research. In addition to the included perceptual and acoustic parameters indicating vocal quality, vocal analysis should also integrate other parameters, such as prosodic (e.g. speech rate, pause time), spectral and cepstral analyses, to introduce parameters with different properties. One limitation of the current study is the exclusive use of a sustained vowel phonation task, which, although widely used in acoustic analysis, may not fully capture the variability in prosodic features typically observed in continuous speech. Given that different voice tasks were selected based on their suitability for acoustic and perceptual analyses, future studies may assess whether the findings remain consistent across different tasks and analyses, including comparisons of the same task evaluated through both acoustic and perceptual methods.

Future research should focus on comprehensive vocal analysis using a large sample of participants, incorporating a wider range of parameters and diverse speech tasks (e. g., reading, sustained vowel, continuous speech) to evaluate the consistency and generalizability of prediction results. It would also be important to compare these results with findings from studies in other languages. Furthermore, the effect of medication on the voice should be explored, along with smoking and coffee consumption, which may alter the therapeutic effects of medication (Radmanović et al., 2017). In future research, participants should be followed longitudinally to monitor voice characteristics across clinically relevant stages, from diagnosis and treatment response to relapse. It would be significant to identify causal factors associated with voice characteristics specific to depression. Since depression is associated with heterogeneous factors, it would be significant to examine the role of individual factors to assess the contribution of intraindividual factors to the voice. Given that psychomotor slowing and agitation may have opposing effects on speech and voice characteristics, future studies should consider examining their individual contributions, potentially by dividing participants into subgroups based on dominant symptoms. Therefore, future research should move towards creating more complex machine-learning models and neural networks that determine both inter- and intraindividual differences to deepen these knowledge. These models should take into account sample size, demographic characteristics, languages, analyzed parameters, speech tasks, and depression scale assessments when creating algorithms. Moreover, fostering multidisciplinary collaboration among psychiatrists, speech therapists, psychologists, and AI engineers could be important to better harness the potential of voice analysis in depression. Such advances may help standardize vocal analysis in depression and enable automatic voice recognition systems to serve as interdisciplinary tools supporting early diagnosis and treatment monitoring.

Conclusion

This study represents the first known attempt to identify depression severity predictors based on voice characteristics in the Serbian-speaking area. Hierarchical regression analysis shows that the acoustic parameter of amplitude peak variation (vAm) and perceptual parameters of hoarseness, breathiness, asthenia, and strain have significant predictive value in determining depression severity. These preliminary findings indicate that voice characteristics hold promise for predicting depression severity (MADRS score). Further research is needed to address the limitations of this study and to ensure generalizability. The obtained results support the potential incorporation of both perceptual and acoustic characteristics (specifically from the domain of intensity variability) within a depression recognition model. If confirmed in larger samples and with more rigorous methodologies, such a model could have important diagnostic and therapeutic implications in clinical practice.

Acknowledgements

The sample and data in this paper are part of the doctoral thesis titled “Impact of voice characteristics on quality of communication in adults with depressive disorders” by Gordana Calić. The study was supported by the Ethics Committee of the University Clinical Center Kragujevac, Serbia (no. 01/21-422). The authors would like to express their gratitude to all the participants who took part in the study.

Conflict of interests

The authors declare no conflict of interest.

Author Contributions

Conceptualization, G.C., B.R., M.P.L., D.I.R., N.S. and M.M.; methodology, G.C.; investigation, G.C. and B.R.; software, M.P.L.; formal analysis, G.C.; writing—original draft preparation, G.C. and B.R.; writing—review and editing, G.C., B.R., M.P.L., D.I.R, N.S and M.M. All authors have read and agreed to the published version of the manuscript.

References

Abitbol, J., Abitbol, P., & Abitbol, B. (1999). Sex hormones and the female voice. *Journal of Voice*, *13*(3), 424-446. https://doi.org/10.1016/S0892-1997(99)80048-4

Afshan, A., Guo, J., Park, S. J., Ravi, V., Flint, J., & Alwan, A. (2018). Effectiveness of voice quality features in detecting depression. *Proceedings of the 19th Annual Conference of the International Speech Communication Association (Interspeech 2018)*, 1676-1680. https://doi.org/10.21437/Interspeech.2018-1399

Alghowinem, S., Goecke, R., Wagner, M., Epps, J., Parker, G., & Breakspear, M. (2013). Characterising depressed speech for classification. *Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech 2013)*, 2534-2538. https://doi.org/10.21437/Interspeech.2013-571

Almaghrabi, S. A., Clark, S. R., & Baumert, M. (2023). Bio-acoustic features of depression: A review. *Biomedical Signal Processing and Control*, *85*, 105020. https://doi.org/10.1016/j.bspc.2023.105020

Alpert, M., Pouget, E. R., & Silva, R. R. (2001). Reflections of depression in acoustic measures of the patient's speech. *Journal of Affective Disorders*, *66*(1), 59-69. https://doi.org/10.1016/S0165-0327(00)00335-9

American Psychiatric Association (APA) (2013). *Diagnostic and Statistical Manual of Mental Disorders (5th ed.)*. https://doi.org/10.1176/appi.books.9780890425596

Arsenić, I., Jovanović Simić, N., Petrović Lazić, M., Šehović, I., & Drljan, B. (2021). Characteristics of speech and voice as predictors of the quality of communication in adults with hypokinetic dysarthria. *Serbian Journal of Experimental and Clinical Research*, *22*(2), 157-165. https://doi.org/10.2478/sjecr-2018-0081

Ayoub, M. R., Larrouy-Maestri, P., & Morsomme, D. (2019). The effect of smoking on the fundamental frequency of the speaking voice. *Journal of Voice*, *33*(5), 802.e11-802.e16. https://doi.org/10.1016/j.jvoice.2018.04.001

Bjelica, M. (2012). *Speech rhythm in English and Serbian: A critical study of traditional and modern approaches*. Filozofski fakultet Novi Sad. ISBN 978-86-6065-111-4

Calić, G., Petrović-Lazić, M., Mentus, T., & Babac, S. (2022a). Akustičke karakteristike glasa kod odraslih osoba sa depresivnim poremećajem. *Psihološka istraživanja*, *25*(2), 183-203. https://doi.org/10.5937/psistra25-39224

Calić, G., Glumbić, N., Petrović-Lazić, M., Đorđević, M., & Mentus, T. (2022b). Searching for best predictors of paralinguistic comprehension and production of emotions in communication in adults with moderate intellectual disability. *Frontiers in Psychology*, *13*, 884242. https://doi.org/10.3389/fpsyg.2022.884242

Cannizzaro, M., Harel, B., Reilly, N., Chappell, P., & Snyder, P. J. (2004). Voice acoustical measurement of the severity of major depression. *Brain and Cognition*, *56*(1), 30-35. https://doi.org/10.1016/j.bandc.2004.05.003

Chlasta, K., Wołk, K., & Krejtz, I. (2019). Automated speech-based screening of depression using deep convolutional neural networks. *Procedia Computer Science*, *164*, 618-628. https://doi.org/10.1016/j.procs.2019.12.228

Ćuk-Jovanović, L. (2002). Akustička analiza govornog signala pacijenata sa depresivnim poremećajem – karakteristike trajanja. *Engrami*, *24*(2), 15-23.

Ćuk-Jovanović, L. (2003). Intenzitet govornog signala pacijenata sa depresivnim poremećajem. *Govor i jezik* (str. 217-223). Institut za eksperimentalnu fonetiku i patologiju govora. ISBN 86-81879-06-5

Cummins, N., Sethu, V., Epps, J., Schnieder, S., & Krajewski, J. (2015). Analysis of acoustic space variability in speech affected by depression. *Speech Communication*, *75*, 27-49. https://doi.org/10.1016/j.specom.2015.09.003

Cummins, N., Sethu, V., Epps, J., Williamson, J. R., Quatieri, T. F., & Krajewski, J. (2020). Generalized two-stage rank regression framework for depression score prediction from speech. *IEEE Transactions on Affective Computing*, *11*(2), 272-283. https://doi.org/10.1109/TAFFC.2017.2766145

Darby, J. K., Simmons, N., & Berger, P. A. (1984). Speech and voice parameters of depression: A pilot study. *Journal of Communication Disorders*, *17*(2), 75-85. https://doi.org/10.1016/0021-9924(84)90013-3

Du, M., Zhang, W., Wang, T., Liu, S., & Ming, D. (2022). An automatic depression recognition method from spontaneous pronunciation using machine learning. *Proceedings of the 2022 9th International Conference on Biomedical and Bioinformatics Engineering (ICBBE '22)*, 133-139. https://doi.org/10.1145/3574198.3574219

Ellgring, H., & Scherer, K. R. (1996). Vocal indicators of mood change in depression. *Journal of Nonverbal Behavior*, *20*(2), 83-110. https://doi.org/10.1007/BF02253071

Gerratt, B. R., Kreiman, J., & Garellek, M. (2016). Comparing measures of voice quality from sustained phonation and continuous speech. *Journal of Speech Language and Hearing Research*, *59*(5), 994-1001. https://doi.org/10.1044/2016_JSLHR-S-15-0307

Hashim, N. W., Wilkes, M., Salomon, R., Meggs, J., & France, D. J. (2017). Evaluation of voice acoustics as predictors of clinical depression scores. *Journal of Voice*, *31*(2), 256.e1-256.e6. https://doi.org/10.1016/j.jvoice.2016.06.006

Huang, X., Wang, F., Gao, Y., Liao, Y., Zhang, W., Zhang, L., & Xu, Z. (2024). Depression recognition using voice-based pretraining model. *Scientific Reports*, *14*, 12734. https://doi.org/10.1038/s41598-024-63556-0

Hönig, F., Batliner, A., Nöth, E., Schnieder, S., & Krajewski, J. (2014). Automatic modelling of depressed speech: Relevant features and relevance of gender. *Proceedings of the 15th Annual Conference of the International Speech Communication Association (INTERSPEECH 2014),* 1248-1252. https://opus.bibliothek.uni-augsburg.de/opus4/frontdoor/deliver/index/docId/67964/file/i14_1248.pdf

Ignjatović Ristić D., Hinić, D., & Jović, J. (2012). Evaluation of the Beck Depression Inventory in a nonclinical student sample. *West Indian Medical Journal*, *61*(5), 489-493. https://scidar.kg.ac.rs/handle/123456789/9559

Isshiki, N., Okamura, H., Tanabe, M., & Morimoto, M. (1969). Differential diagnosis of hoarseness. *Folia Phoniatrica et Logopaedica*, *21*(1), 9-19. https://doi.org/10.1159/000263230

Jia, Y., Liang, Y., & Zhu, T. (2019). An analysis of voice quality of Chinese patients with depression. *Proceedings of the 22nd Conference of the Oriental COCOSDA International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA)*, 1-6. https://doi.org/10.1109/O-COCOSDA46868.2019.9060848

Jiang, H., Hu, B., Liu, Z., Yan, L., Wang, T., Liu, F., Kang, H., & Li, X. (2017). Investigation of different speech types and emotions for detecting depression using different classifiers. *Speech Communication*, *90*, 39-46. https://doi.org/10.1016/j.specom.2017.04.001

Kiss, G., & Jenei, A. Z. (2020). Investigation of the accuracy of depression prediction based on speech processing. *Proceedings of the 43rd International Conference on Telecommunications and Signal Processing (TSP)*, 129-132. https://doi.org/10.1109/TSP49548.2020.9163495

Lasser, K., Wesley, B. J., Steffie, W., Himmelstein, D. U., McCormick, D., & Bor, D. H. (2000). Smoking and mental illness. *JAMA*, *284*(20), 2606-2610. https://doi.org/10.1001/jama.284.20.2606

Laukkanen, A-M., & Sundberg, J. (2008). Peak-to-peak glottal flow amplitude as a function of F0. *Journal of Voice*, *22*(6), 614-621. https://doi.org/10.1016/j.jvoice.2007.01.003

Liang, L., Wang, Y., Ma, H., Zhang, R., Liu, R., Zhu, R., Zheng, Z., Zhang, X., & Wang, F. (2024). Enhanced classification and severity prediction of major depressive disorder using acoustic features and machine learning. *Frontiers in Psychiatry*, *15*, 1422020. https://doi.org/10.3389/fpsyt.2024.1422020

Liu, Z., Hu, B., Liu, F., & Kang, H. (2016). Evaluation of depression severity in speech. *Proceedings of the International Conference on Brain and Health Informatics (BHI 2016)*, 312–321. https://doi.org/10.1007/978-3-319-47103-7_31

Menne, F., Dörr, F., Schräder, J., Tröger, J., Habel, U., König, A., & Wagels, L. (2024). The voice of depression: speech features as biomarkers for major depressive disorder. *BMC Psychiatry*, *24*(1), 794. https://doi.org/10.1186/s12888-024-06253-6

Mihajlović, G., Vojvodić, P., Vojvodić, J., Andonov, A., & Hinić, D. (2021). Validation of the Montgomery-Åsberg depression rating scale in depressed patients in Serbia. *Srpski arhiv za celokupno lekarstvo*, *149*(5-6), 316-321. https://doi.org/10.2298/SARH200401004M

Montgomery, S. A., & Åsberg, M. (1979). A new depression scale designed to be sensitive to change. *The British Journal of Psychiatry*, *134*, 382-389. https://doi.org/10.1192/bjp.134.4.382

Mundt, J. C., Snyder, P. J., Cannizzaro, M. S., Chappie, K., & Geralts, D. S. (2007). Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. *Journal of Neurolinguistics, 20*(1), 50-64. https://doi.org/10.1016/j.jneuroling.2006.04.001

Mundt, J. C., Vogel, A. P., Feltner, D. E., & Lenderking, W. R. (2012). Vocal acoustic biomarkers of depression severity and treatment response. *Biological Psychiatry*, *72*(7), 580-587. https://doi.org/10.1016/j.biopsych.2012.03.015

Müller, M. J., Himmerich, H., Kienzle, B., & Szegedi, A. (2003). Differentiating moderate and severe depression using the Montgomery–Åsberg depression rating scale (MADRS). *Journal of Affective Disorders*, *77*(3), 255-260. https://doi.org/10.1016/s0165-0327(02)00120-9

Nejati, S., Ariai, N., Björkelund, C., Skoglund, I., Petersson, E-L., Augustsson, P., Hange, D., & Svenningsson, I. (2020). Correspondence between the Neuropsychiatric Interview M.I.N.I. and the BDI-II and MADRS-S self-rating instruments as diagnostic tools in primary care patients with depression. *International Journal of General Medicine*, *13*, 177-183.

Nguyen, D. D., Novakovic, D., & Madill, C. (2024). Voice disorder discrimination using vowel acoustic measures in female speakers. *International Journal of Language & Communication Disorders*, *59*(5), 2087-2102. https://doi.org/10.1111/1460-6984.13081

Nikolić, D. (2016). Acoustic analysis of English vowels produced by American speakers and highly competent Serbian L2 speakers. *Facta Universitatis Series: Linguistics and Literature*, *14*(1), 85-101.

Petrović-Lazić, M., & Kosanović, R. (2008). *Vokalna rehabilitacija glasa*. Nova naučna. ISBN 978-86-87449-00-8

Petrović-Lazić, M., Jovanović-Simić, N., Šehović, I., & Ćalasan, S. (2016). Uticaj zamora na akustičke karakteristike glasa kod vokalnih profesionalaca, *Biomedicinska istraživanja*, *7*(1), 6-10. https://doi.org/10.7251/BII1601006P

Petrović-Lazić, M. (2021). *Instrumentalne i test metode kliničkog ispitivanja glasa*. Nova poetika. ISBN 978-86-902700-2-6

Petrović-Lazić, M., & Ilić Savić, I. (2023). Changes in the level of sex hormones with aging and their influence on the voice. *Zdravstvena zaštita*, *52*(3), 56-65. https://www.doi.org/10.5937/zdravzast52-44412

Quatieri, T., & Malyska, N. (2012). Vocal-source biomarkers for depression: A link to psychomotor activity. *Proceedings of the 13th Annual Conference of the International Speech Communication Association (Interspeech 2012)*, 1059-1062. https://www.isca-archive.org/interspeech_2012/quatieri12_interspeech.pdf

Radmanović, B., Đukić-Dejanović, S., Milovanović, D. R., & Đorđević, N. (2017). Cigarette smoking and heavy coffee drinking affect therapeutic response to olanzapine. *Srpski arhiv za celokupno lekarstvo*, *146*(1-2), 43-47. https://doi.org/10.2298/SARH170307122R

Rejaibi, E., Komaty, A., Meriaudeau, F., Agrebi, S., & Othmani, A. (2022). MFCC-based Recurrent Neural Network for automatic clinical depression recognition and assessment from speech. *Biomedical Signal Processing and Control*, *71*, 103107. http://doi.org/10.1016/j.bspc.2021.103107

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*, 206-215. https://doi.org/10.1038/s42256-019-0048-x

Sahu, S., & Espy-Wilson, C. (2016). Speech features for depression detection. *Proceedings of the 17th Annual Conference of the International Speech Communication Association (Interspeech 2016)*, 1928-1932. https://www.isca-archive.org/interspeech_2016/sahu16_interspeech.pdf

Šehović, I., Petrović-Lazić, M., & Jovanović-Simić, N. (2017). Akustička i perceptivna analiza ezofagealnog i traheoezofagealnog glasa. *Specijalna edukacija i rehabilitacija*, *16*(3), 289-307. https://doi.org/10.5937/specedreh16-13683

Seneviratne, N., & Espy-Wilson, C. (2021). Speech based depression severity level classification using a multi-stage dilated CNN-LSTM model. *Proceedings of the 22nd Annual Conference of the International Speech Communication Association (Interspeech 2021)*, 2526-2530. https://doi.org/10.21437/Interspeech.2021-1967

Shin, D., Cho, W. I., Park, C. H. K., Rhee, S. J., Kim, M. J., Lee, H., Kim, N. S., & Ahn, Y. M. (2021). Detection of minor and major depression through voice as a biomarker using machine learning. *Journal of Clinical Medicine*, *10*(14), 3046. https://doi.org/10.3390/jcm10143046

Silva, W. J., Lopes, L., Galdino, M. K. C., & Almeida, A. A. (2024). Voice acoustic parameters as predictors of depression. *Journal of Voice*, 38(1), 77-85. https://doi.org/10.1016/j.jvoice.2021.06.018

Songur, E. T., Hafizoğlu, M., Aydinli, F. E., İncebay, Ö, Parlak, M. M., & Balci, C. (2025). Analysis of the auditory-perceptual voice quality in older and younger adults without self-reported voice complaints. *Journal of Voice*, In Press. https://doi.org/10.1016/j.jvoice.2024.12.022

Stráník, A., Čmejla, R., & Vokřál, J. (2014). Acoustic parameters for classification of breathiness in continuous speech according to the GRBAS scale. *Journal of Voice*, *28*(5), 653.e9–653.e17. https://doi.org/10.1016/j.jvoice.2013.07.016

Stubbs, B., Vancampfort, D., Firth, J., Solmi, M., Siddiqi, N., Smith, L., Carvalho, A. F., & Koyanagi, A. (2018). Association between depression and smoking: A global perspective from 48 low- and middle-income countries. *Journal of Psychiatric Research*, *103*, 142-149. https://doi.org/10.1016/j.jpsychires.2018.05.018

Taguchi, T., Tachikawa, H., Nemoto, K., Suzuki, M., Nagano, T., Tachibana, R., Nishimura, M., & Arai, T. (2017). Major depressive disorder discrimination using vocal acoustic features. *Journal of Affective Disorders*, *225*, 214-220. https://doi.org/10.1016/j.jad.2017.08.038

Vahid-Ansari, F. & Albert, P. R. (2021). Rewiring of the serotonin system in major depression. *Frontiers in Psychiatry*, *12*, 802581. https://doi.org/10.3389/fpsyt.2021.802581

Wadle, L. M., Ebner-Priemer, U. W., Foo, J. C., Yamamoto, Y., Streit, F., Witt, S. H., Frank, J., Zillich, L., Limberger, M. F., Ablimit, A., Schultz, T., Gilles, M., Rietschel, M., & Sirignano, L. (2024). Speech features as predictors of momentary depression severity in patients with depressive disorder undergoing sleep deprivation therapy: Ambulatory assessment pilot study. *JMIR Mental Health*, *11*, e49222. https://doi.org/10.2196/49222.

Wang, J., Zhang, L., Liu, T., Pan, W., Hu, B., & Zhu, T. (2019). Acoustic differences between healthy and depressed people: a cross-situation study. *BMC Psychiatry*, *19*(1), 300. https://doi.org/10.1186/s12888-019-2300-7

Wang, Y., Liang, L., Zhang, Z., Xu, X., Liu, R., Fang, H., Zhang, R., Wei, Y., Liu, Z., Zhu, R., Zhang, X., & Wang, F. (2023). Fast and accurate assessment of depression based on voice acoustic features: a cross-sectional and longitudinal study. *Frontiers in Psychiatry*, *14*, 1195276. https://doi.org/10.3389/fpsyt.2023.1195276

Williamson, J. R., Young, D., Nierenberg, A. A., Niemi, J., Helfer, B. S., & Quatieri, T. F. (2018). Tracking depression severity from audio and video based on speech articulatory coordination. *Computer Speech & Language*, *55*, 40-56. https://doi.org/10.1016/j.csl.2018.08.004

Yalamanchili, B., Kota, N. S., Abbaraju, M. S., Nadella, V. S. S., & Alluri, S. V. (2020). Real-time acoustic based depression detection using machine learning techniques. *Proceedings of the 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, 1-6. https://ieeexplore.ieee.org/document/9077698

Yamamoto, M., Takamiya, A., Sawada, K., Yoshimura, M., Kitazawa, M., Liang, K-C., Fujita, T., Mimura, M., & Kishimoto, T.

(2020). Using speech recognition technology to investigate the association between timing-related speech features and depression severity. *PLoS ONE*, *15*(9), e0238726. https://doi.org/10.1371/journal.pone.0238726

Yang, Y., Fairbairn, C., & Cohn, J. F. (2013). Detecting depression severity from vocal prosody. *IEEE Transactions on Affective Computing*, *4*(2), 142-150. https://doi.org/10.1109/T-AFFC.2012.38

Yu, Y. H., Shafer, V. L., & Sussman, E. S. (2017) Neurophysiological and behavioral responses of Mandarin lexical tone processing. *Frontiers in Neuroscience*, *11*, 95. https://doi.org/10.3389/fnins.2017.00095

Zhang, L., Duvvuri, R., Chandra, K. K. L., Nguyen, T., & Ghomi, R. H. (2020). Automated voice biomarkers for depression symptoms using an online cross-sectional data collection initiative. *Depression and Anxiety*, *37*(7), 657-669. https://doi.org/10.1002/da.23020

Zhao, Q., Fan, H-Z., Li, Y-L., Liu, L., Wu, Y-X., Zhao, Y-L., Tian, Z-X., Wang, Z-R., Tan, Y-L., & Tan, S-P. (2022). Vocal acoustic features as potential biomarkers for identifying/diagnosing depression: A cross-sectional study. *Frontiers in Psychiatry*, *13*, 815678. https://doi.org/10.3389/fpsyt.2022.815678